

Continuous-time Proportional-Integral Distributed Optimization for Networked Systems

Greg Droge^{a1} and Hiroaki Kawashima^b and Magnus Egerstedt^a

^aSchool of Electrical and Computer Engineering, Georgia Institute of Technology, Atlanta, GA 30332, USA.

^bGraduate School of Informatics, Kyoto University, Kyoto 606-8501, Japan.

Email: gregdroge@gatech.edu, kawashima@i.kyoto-u.ac.jp, magnus@gatech.edu

Phone: +1-404-385-4077, +81-75-753-3327, +1-404-894-3484

The work by M. Egerstedt was funded by The Air Force Office of Scientific Research through grant number [2012-00305-01]

¹To whom correspondence should be addressed.

Continuous-time Proportional-Integral Distributed Optimization for Networked Systems

Greg Droge and Hiroaki Kawashima and Magnus Egerstedt

Abstract

In this paper we explore the relationship between dual decomposition and the consensus-based method for distributed optimization. The relationship is developed by examining the similarities between the two approaches and their relationship to gradient-based constrained optimization. By formulating each algorithm in continuous-time, it is seen that both approaches use a gradient method for optimization with one using a proportional control term and the other using an integral control term to drive the system to the constraint set. Therefore, a significant contribution of this paper is to combine these methods to develop a continuous-time proportional-integral distributed optimization method. Furthermore, we establish convergence using Lyapunov stability techniques and utilizing properties from the network structure of the multi-agent system.

Keywords

Distributed Optimization, Multi-agent control, Networked-systems

1 Introduction

In recent years, much attention has been given to the control, optimization, and coordination of multi-agent systems. Multi-agent systems present many challenging problems, while having applications as diverse as data fusion in sensor networks to multiple robots acting in collaboration, e.g. (Olfati-Saber, 2007; Mesbahi & Egerstedt, 2010; Shamma, 2007; Olfati-Saber & Murray, 2004; Jadbabaie, Lin, & Morse, 2003). A significant amount of research has been produced on the design of algorithms to allow agents to collaborate and achieve desirable global properties in a distributed manner, e.g. (Mesbahi & Egerstedt, 2010; Shamma, 2007; Palomar & Eldar, 2010). In this paper, we will address the problem of distributed optimization, where agents in the network collaborate to minimize a global cost and each agent uses only locally available information, defined by the network structure.

While many classifications of distributed optimization algorithms may exist, a distinction has recently been made in both (Wei & Ozdaglar, 2012) and (Kvaternik & Pavel, 2012) that distributed optimization techniques can be divided between two categories: consensus-based gradient methods and decomposition or primal-dual methods. The consensus-based approach is characterized by algorithms where, at each time step, every agent takes a gradient step along with an averaging or consensus step to reach agreement on the variables being optimized, e.g. (Kvaternik & Pavel, 2012; Wang

& Elia, 2010; Nedic & Ozdaglar, 2009a; Wang & Elia, 2011). In contrast, decomposition methods distributedly reach agreement by exploiting the dual of the problem, e.g. (Terelius, Topcu, & Murray, 2011; Rantzer, 2007, 2009), which requires the added collaborative update of pricing or dual variables, e.g. (Nedic & Ozdaglar, 2009b; Luenberger & Ye, 2008; Boyd, 2004). Of particular interest is the decomposition method for multi-agent systems presented in (Terelius et al., 2011) along with the gradient-based solution for dual problems first presented in (Arrow, Hurwicz, & Uzawa, 1958). When combined, these methods allow for a gradient-based multi-agent distributed optimization technique that, for simplicity, we refer to as dual-decomposition.

In this paper, we show that the consensus-based and dual-decomposition gradient algorithms are actually very closely related when examined in context of the underlying constrained optimization problem that is solved by these methods. Specifically, we formulate both the mentioned dual-decomposition method and the consensus-based method in (Kvaternik & Pavel, 2012) in control theoretic terms to draw parallels and gain intuition behind why they can naturally be joined together. In fact, it will become apparent that dual-decomposition is very closely related to integral (I) control and the consensus method is closely related to proportional (P) control. Therefore, a significant contribution of this paper is to combine these two methods to form a new, proportional-integral (PI) distributed optimization method. This formulation will be similar to the PI distributed optimization method introduced in (Wang & Elia, 2010) and extended in (Wang & Elia, 2011; Gharesifard & Cortés, 2012). However, due to the fact that we create the PI optimization method from the perspective of the dual-decomposition method, which involves a set of constraints associated with the interconnections of agents, it enables us to form a different type of integral control term and reduce the required communication.

While much of the work on distributed optimization has been developed in discrete-time formulations, which are amenable for implementation, e.g. (Wei & Ozdaglar, 2012; Nedic & Ozdaglar, 2009a; Terelius et al., 2011; Lobel & Ozdaglar, 2008), a great deal of work recently has been made in continuous-time (Kvaternik & Pavel, 2012; Wang & Elia, 2010, 2011; Rantzer, 2007, 2009; Kvaternik & Pavel, 2011; Gharesifard & Cortés, 2012). Continuous-time analysis has proven useful as it allows Lyapunov stability conditions to be directly applied to the update-equations for convergence analysis. It also allows for an intuitive connection between the optimization algorithm proposed in this paper and proportional-integral control. Moreover, a discretization of the framework proposed in this paper does not pose a significant contribution. The proportional element has been evaluated in discrete time in (Nedic & Ozdaglar, 2009a) and the integral element has been evaluated in discrete time in

(Terelius et al., 2011; Rantzer, 2007, 2009). Furthermore, a closely related PI distributed optimization algorithm developed in (Wang & Elia, 2010, 2011; Gharesifard & Cortés, 2012) (discussed further in Section 5.1) was discretized in (Wang & Elia, 2010).

The remainder of this paper will proceed as follows. We first introduce the necessary background material for the analysis of the distributed optimization algorithms, including the problem formulation, the graph-based multi-agent model of the network, and a parallel to gradient-based constrained optimization. This background will allow us to present both dual-decomposition and the consensus based method in Sections 3 and 4. These two methods are then combined in Section 5 to create a PI distributed optimization method. We further develop this method by presenting a formulation which is scalable to a larger class of multi-agent systems in Section 6. The paper is then concluded with some final remarks in Section 7.

2 Preliminaries

This section introduces the background information necessary to characterize PI distributed optimization. It begins with the formulation of the distributed optimization problem that is addressed in this paper. Following this, the graph-based model of the multi-agent network will be introduced. Gradient-based constrained optimization is then discussed from a high level viewpoint to develop intuition about the underlying relationship between dual-decomposition and the consensus based method. As similarities to PI control will become readily apparent, this section ends with a brief introduction of the PI control metrics that are used to compare these methods.

2.1 Problem formulation

We address the distributed optimization methods in terms of the problem formulation presented in (Nedic & Ozdaglar, 2009a) and continued in (Kvaternik & Pavel, 2012; Wang & Elia, 2010, 2011; Sundhar Ram, Nedić, & Veeravalli, 2010; Gharesifard & Cortés, 2012; Kvaternik & Pavel, 2011). Specifically, assume that the function to be optimized is a summation of strictly-convex costs, i.e.

$$\min_x \sum_{i=1}^N f_i(x), \quad (1)$$

where $x \in \mathbb{R}^n$ is the parameter vector being optimized, N is the number of agents, and agent i only knows its individual cost, $f_i(x)$. The individual costs can be derived naturally from a distributed

problem as done in (Palomar & Eldar, 2010) for resource allocation, or can be “designed” as done, for example, in (Keviczky, Borrelli, & Balas, 2006; Droge & Egerstedt, 2013), where a central cost is split into separable components which are then assigned to the individual agents.

To be able to establish convergence to the global minimum certain convexity assumptions are made on the cost. Note that a convex function is defined as a function that satisfies:

$$f(\theta x + (1 - \theta)y) \leq \theta f(x) + (1 - \theta)f(y) \quad (2)$$

where $0 < \theta < 1$. A function is strictly-convex if strict inequality holds in (2) (see, for example, (Boyd, 2004) for a thorough overview of convex functions and their properties). The following assumptions about the costs are used throughout the paper:

Assumption 1. $f_i(x) : \mathbb{R}^n \rightarrow \mathbb{R}$, $i \in \{1, \dots, N\}$, are convex, twice continuously differentiable functions and the summation $\sum_{i=1}^N f_i(x)$ is strictly-convex, twice continuously differentiable function.

Assumption 2. The solution $f^* = \min_x \sum_{i=1}^N f_i(x)$ and respective optimal parameter vector, x^* , exist and are finite.

Remark 1. We note that the differentiability assumption has been relaxed in many of the references to address subgradient optimization. However, we do not concern ourselves with relaxing this assumption as it does not add to the development of the paper.

For sake of clarifying the notation, one key point must be stressed. To perform distributed optimization, each agent will maintain its “own version” of the variables, denoted as $x_i \in \mathbb{R}^n$, with the constraint that $x_i = x_j \forall i, j \in \{1, \dots, N\}$. This will allow (1) to be expressed as

$$\begin{aligned} \min_{x_i, i=1, \dots, N} \sum_{i=1}^N f_i(x_i). \\ \text{s.t. } x_i = x_j \forall i, j \in \{1, \dots, N\} \end{aligned} \quad (3)$$

To perform the optimization in a distributed manner, the equality constraints are relaxed. Algorithms differ in the manner that they force agents to return to the constraint set.

2.2 Networked multi-agent systems

We now introduce the terminology and properties of multi-agent systems that will be used to formulate the distributed optimization algorithms and discuss their convergence. The term “agent” is used to

refer to a computing component and it is assumed that agents only communicate with each other through a defined, static network topology. This is representative of a great number of different multi-agent systems, from communication networks to teams of robots, e.g. (Mesbahi & Egerstedt, 2010; Shamma, 2007; Rantzer, 2009).

The interconnections of the network are represented through an undirected graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$. The set of nodes, \mathcal{V} , is defined such that $v_i \in \mathcal{V}$ corresponds to agent i . Communication constraints are represented through the set of edges in the graph, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$, where $(v_i, v_j) \in \mathcal{E}$ iff agents i and j can directly communicate. The number of agents is then given by $|\mathcal{V}| = N$ and the number of communication links is given by $|\mathcal{E}| = M$. To prove convergence of the distributed optimization methods, the following assumption on the graph topology is made

Assumption 3. *The graph $\mathcal{G}(\mathcal{V}, \mathcal{E})$ is connected.*

Associated with this graph are two important, and related matrices. The first is the incidence matrix, $D \in \mathbb{R}^{N \times M}$ which is formed by arbitrarily assigning an orientation to each edge and can be defined as

$$D = [d_{ik}] = \begin{cases} 1 & \text{edge } k \text{ points to node } i \\ -1 & \text{edge } k \text{ originates at node } i \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

The second matrix, the graph Laplacian, is closely related to D and can be defined as $L = L^T = DD^T \in \mathbb{R}^{N \times N}$. Note that the resulting values for the elements of L are independent of the orientation assigned to each edge, (Mesbahi & Egerstedt, 2010).

We utilize both the incidence matrix and the graph Laplacian to form larger, aggregate matrices to incorporate the fact that each agent will be maintaining an entire vector of values. First, let x_{ij} denote the j^{th} element of x_i , $z_j \triangleq [x_{1j}, x_{2j}, \dots, x_{Nj}]^T$ is the combination of all the j^{th} elements, and $z \triangleq [z_1^T, \dots, z_n^T]^T \in \mathbb{R}^{Nn}$ is the aggregate state vector. The aggregate matrices can then be written as $\mathbb{D} \triangleq I_n \otimes D$ and $\mathbb{L} \triangleq I_n \otimes L$. This notation expresses the concept that an aggregate graph is formed where there are n replicas of \mathcal{G} , each corresponding to one of the elements of the vector being optimized. The aggregate graph will not be connected, but have n connected components, given Assumption 3. Therefore, the aggregate Laplacian will have the following properties (see, for example, (Mesbahi & Egerstedt, 2010)):

$$(1) \quad \mathbb{L} = \mathbb{L}^T = \mathbb{D}\mathbb{D}^T$$

(2) $\mathbb{L} \succeq 0$

(3) The eigenvectors associated with the zero eigenvalues of the aggregate Laplacian are $\alpha \otimes \mathbf{1}$, where $\alpha \in \mathbb{R}^n$

(4) If $\dot{z} = -\mathbb{L}z$, the solution, $\bar{z} = z(t)$ as $t \rightarrow \infty$, will be the projection of $z(0)$ onto the set $\alpha \otimes \mathbf{1}$ for $\alpha \in \mathbb{R}^n$. Moreover, the vector $-\mathbb{L}z$ will point along a line orthogonal to the set $\{\alpha \otimes \mathbf{1} | \alpha \in \mathbb{R}^n\}$.

One further property that will be exploited throughout the paper comes from the incidence matrix. The constraint in (3) that $x_i = x_j \forall i, j \in \{1, \dots, N\}$ can be written as $\mathbb{D}^T z = 0$. This can be verified by first considering the scalar case where $n = 1$ and $\mathbb{D} = D$. $D^T z = 0$ will enforce that $x_{k_1} - x_{k_2} = 0$, where k_1 and k_2 correspond to the vertices associated with edge k . Then through Assumption 3, $x_i = x_j \forall i, j \in \{1, \dots, N\}$. The same argument can be extended to $n > 1$ case by noting that:

$$\mathbb{D}^T z = \begin{bmatrix} D^T z_1 \\ \vdots \\ D^T z_n \end{bmatrix}.$$

Finally, this notation allows the distributed optimization problem to be presented in a compact form:

$$\begin{aligned} \min_z f(z) \\ \text{s.t. } h(z) = 0. \end{aligned} \tag{5}$$

where $f(z) = \sum_{i=1}^N f_i(x_i)$ and $h(z) = \mathbb{D}^T z$.

2.3 PI control as gradient method for constrained optimization

We now take note of the structure of (5) to give intuition to the relationship between the gradient methods presented in Sections 3 and 4. The development in this section will not dwell on the details of constrained optimization, as it has been a widely studied area, e.g. (Luenberger & Ye, 2008; Boyd, 2004). Rather, it will be focused on forming a control law to return the state to the constraint set when the constraints are relaxed.

Without constraints, a gradient method for optimization of the problem would simply take the form $\dot{z} = -k_G \frac{\partial f}{\partial z}^T$, where $k_G \in \mathbb{R}^+$ is some gain. However, when the optimization includes constraints,

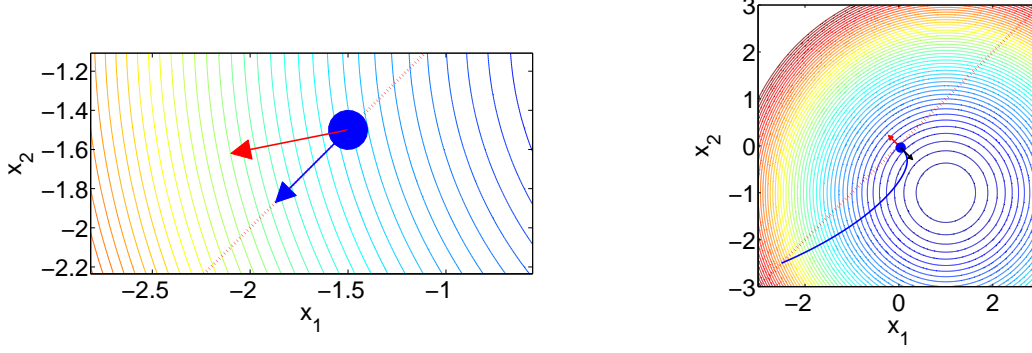


Figure 1: This figure shows the results of using the cost $f(z) = (x_1 - 1)^2 + (x_2 + 1)^2$. Left: Dotted line shows the equality constraint and the arrows show the gradient and projected gradient. Right: Result of performing the PI gradient method for optimization given in (8). The trajectory of the two states is shown ending in the final condition denoted by the solid circle and the constraint is shown as a dotted line. The arrows show the final gradient and Lagrange multiplier multiplied by the constraint. As expected, these are equal in magnitude, but opposite in direction.

the update to the variables being optimized cannot be in any arbitrary direction. The update can only occur in a direction that will allow the state to continue to satisfy the constraint. As the constraints in (5) are linear, this involves taking the gradient and projecting it onto the constraint space, as shown in Figure 1.

It should be noted that the difference between an unconstrained gradient and a constrained gradient could be written in terms of the addition of a term perpendicular to the constraint set. This could be expressed as $\frac{\partial f}{\partial z} + \lambda^T \frac{\partial h}{\partial z} = \frac{\partial f}{\partial z} + \lambda^T \mathbb{D}^T$. The dynamics of the resulting optimization would then be

$$\dot{z} = -k_G \left(\frac{\partial f}{\partial z} + \mathbb{D} \lambda(t) \right). \quad (6)$$

However, computing $\lambda(t)$ in a distributed fashion could be difficult as it may require knowledge from the entire network.

Alternatively, if the gradient method is permitted to violate the constraint, control terms can be added to guide the state back to the constraint at the optimal point. The first term we consider is a term proportional to the error from the constraint. Allow $\lambda(t) = \frac{k_P}{k_G} e(t)$ where $e(t) = \mathbb{D}^T z(t)$ is the error at each edge of the graph. This can be seen to be a logical choice because, as mentioned in Section 2.2, $-\mathbb{D}e(t) = -\mathbb{L}z(t)$ will point along a line orthogonal to the constraint set. In other words, it points in the right direction, but with possibly the wrong magnitude. This gives the dynamics

$$\dot{z} = -k_G \frac{\partial f}{\partial z}(z) - k_P \mathbb{D}e(t) = -k_G \frac{\partial f}{\partial z}(z) - k_P \mathbb{L}z(t). \quad (7)$$

As will be discussed in Section 4, the similarity of (7) to proportional control is perpetuated in that the steady-state solution will have a constant error from the desired optimal point. Basically, the effort produced by introducing an error term proportional to the deviation from the constraint will fall short of the needed effort to drive the state all of the way to the constraint set.

To compensate for the steady state error, it is common to add an integral term to the control, e.g. (Franklin, Powell, & Emami-Naeini, 2001). This would lead to a $\lambda(t)$ of the form $\lambda(t) = \frac{k_P}{k_G}e(t) + \frac{k_I}{k_G} \int_{t_0}^t e(s)ds$. Over time, the integral term will build up the necessary effort to reach the constraint. With this additional term, the dynamics of the system can be expressed as

$$\dot{z} = -k_G \frac{\partial f^T}{\partial z} - \mathbb{D} \left(k_P e(t) + k_I \int_{t_0}^t e(s)ds \right) = -k_G \frac{\partial f^T}{\partial z} - k_P \mathbb{L}^T z - k_I \mathbb{L} \int_{t_0}^t z(s)ds. \quad (8)$$

It will be shown in Section 5 that under Assumptions 1, 2, and 3, the dynamics in (8) will indeed converge to a global minimum, as shown in Figure 1.

While this method for obtaining a gradient strategy to arrive at the desired optimal value may seem somewhat trivial or ad-hoc, it will be seen in Section 3 that the dual-decomposition distributed optimization method will exactly correspond to adding an integral term. Similarly, in Section 4, it is shown that the consensus-based method will be exactly the proportional term. Therefore, we combine the two methods in Section 5 to form a PI distributed optimization method.

2.4 PI performance metrics

As the distributed control laws developed throughout the remainder of this paper are closely related to proportional and integral control laws, we give a brief introduction to the performance metrics that will be employed for comparison. These metrics are important as there really is no single metric which best determines which control law is most suitable. For example, as discussed in (Franklin et al., 2001), proportional control can converge quickly, but may result in a steady-state error. As the proportional gain is increased, the steady-state error will typically decrease up to the point where the system becomes unstable. On the other hand, integral control can be introduced to eliminate steady-state error, but dampening will be decreased and this will result in greater oscillation, overshoot, and slower convergence.

Therefore, to say that one method is “better” would require a reference to a specific application. To be able to judge which method is more suitable for the given application, the following performance metrics, typical for classic control evaluation (e.g. (Franklin et al., 2001)), are used:

- Percent overshoot (M_p): The percentage of the distance that the state overshoots the final value, given as $\frac{x_{max}-x_f}{x_f-x_0} \times 100$.
- Settling time (t_{10} and t_1): Time it takes for the state to converge to within 10 percent and 1 percent of the final value. For example, t_{10} is the smallest t such that $.9\frac{x_f}{x_f-x_0} < x(t) < 1.1\frac{x_f}{x_f-x_0} \forall t > t_{10}$.
- Percent error (% error): The percentage of error from the optimal value ($\frac{|x^*-x_f|}{x_f-x_0} \times 100$).

where x_0 is the initial value, x_f is the final value, and x_{max} is the maximum value reached. For simplicity, we have assumed $x_{max} \geq x_f > x_0$. As these values are measures of scalar states, the worst case over all agents will be presented in each evaluation.

Also note that numerical results depend upon the value of the gains and initial conditions. To allow for a fair comparison between examples throughout the paper, all gains (k_G , k_P , and k_I) are assigned a value of 1. Similarly, all initial conditions are assigned a value of 0, unless otherwise stated.

3 Dual decomposition

This section introduces the concept of gradient-based distributed optimization through the introduction of dual-decomposition, which has been used in a variety of different applications, e.g. (Palomar & Eldar, 2010; Wang & Elia, 2010; Terelius et al., 2011; Droge & Egerstedt, 2013; Rantzer, 2009; Giselsson & Rantzer, 2010). Notation, examples, and proofs are given which will allow for a concise development of the distributed optimization methods in Sections 4 and 5.

As already mentioned, dual-decomposition will be akin to integral control for constrained optimization. However, to provide intuition as to the origins and the theoretical underpinnings of this method, it is presented here in a more typical fashion relying upon the theory of dual-optimization, e.g. (Luenberger & Ye, 2008; Boyd, 2004). The formulation introduced here is closely related to that found in (Terelius et al., 2011), except that we use Uzawa's saddle point method, (Arrow et al., 1958), to update both the parameters and dual variables simultaneously. This permits a continuous-time formulation where Lyapunov methods can be readily applied to establish convergence. After presenting the algorithm, the relation to integral control will be evaluated. This section will end with a distributed implementation and a numerical example.

3.1 Dual-decomposition for networked systems

The basic idea behind dual decomposition is to introduce n copies of the variables, with the constraint that the copies be equal. The dual is then formed to relax the added constraints and a max min optimization technique is used to solve the dual problem. In this paper, we use a gradient method introduced in (Arrow et al., 1958) for saddle point finding. This will allow for a distributed solution to the problem where each agent uses only local information defined by the network graph, \mathcal{G} .

The dual problem to (3) can be formed by introducing a Lagrange multiplier vector, $\mu_k \in \mathbb{R}^n$, $k = 1, \dots, M$, for each edge in \mathcal{G} . It can be written as:

$$\max_{\mu_k, k=1, \dots, M} \min_{x_i, i=1, \dots, N} \left\{ k_G \sum_{i=1}^N f_i(x_i) + k'_I \sum_{k=1}^M \mu_k^T (x_{k_1} - x_{k_2}) \right\} \quad (9)$$

where, again, the subscripts k_1 and k_2 correspond to the agents which make up the k^{th} edge and $k_G, k'_I > 0$ are constant gains. Note that due to the constraint equaling zero, k'_I has no influence and k_G scales the cost, but does not change the location of the optimal point. Equation (9) can be simplified by forming an aggregate Lagrange multiplier vector, $\mu \in \mathbb{R}^{Mn}$, in the same fashion that the aggregate state, z , was formed. This allows us to reintroduce the constraint as $\mathbb{D}^T z = 0$ and rewrite (9) as:

$$\max_{\mu} \min_z F(z, \mu) = k_G f(z) + k'_I z^T \mathbb{D} \mu. \quad (10)$$

To solve this max-min problem, we use a technique first developed in (Arrow et al., 1958) for saddle point finding and has more recently gained attention for its applicability to distributed optimization, e.g. (Wang & Elia, 2010, 2011; Rantzer, 2009, 2007). The basic idea behind this approach is that dynamics can be assigned to the variables being optimized and convergence can be established using control methods such as Lyapunov stability.

For a saddle point finding problem, where $F(z, \mu)$ is strictly-convex in z and strictly concave in μ , (Arrow et al., 1958) shows that applying the dynamics

$$\dot{z} = -\frac{\partial F^T}{\partial z}, \quad \dot{\mu} = \frac{\partial F^T}{\partial \mu}, \quad (11)$$

the system will converge asymptotically to the saddle point. Taking the partials of (10), the dynamics can be expressed as:

$$\dot{z} = -k_G \frac{\partial f^T}{\partial z} - k'_I \mathbb{D} \mu \quad (12)$$

$$\dot{\mu} = k'_I \mathbb{D}^T z. \quad (13)$$

However, we note that (10) is not strictly concave in μ , rather, it is linear. This requires further evaluation, which is done in the proofs of Theorems 1 and 2. While there exist proofs for dual-decomposition, e.g. (Palomar & Eldar, 2010; Feijer & Paganini, 2010), we present an alternative proof here to show the relationship of dual-decomposition to the underlying constrained optimization problem. This will allow us to easily extend these proofs in Section 5 for the PI distributed optimization method that will be developed. The proofs use the same Lyapunov candidate function as (Rantzer, 2007; Feijer & Paganini, 2010) to prove convergence, but differ in the application of Lasalle's invariance principal and the proof that the equilibrium reached is the global minimum.

Theorem 1. *Given Assumptions 1, 2, and 3 as well as the dynamics in (12) and (13), the saddle point $(\dot{z}, \dot{\mu}) = (0, 0)$ is globally asymptotically stable.*

Proof. Using the candidate Lyapunov function $V = \frac{1}{2}(\dot{z}^T \dot{z} + \dot{\mu}^T \dot{\mu})$, \dot{V} can be written as:

$$\begin{aligned} \dot{V} &= \dot{z}^T \ddot{z} + \dot{\mu}^T \ddot{\mu} = -\dot{z}^T \left(k_G \frac{\partial^2 f}{\partial z^2} \dot{z} + k'_I \mathbb{D} \dot{\mu} \right) + k'_I \dot{\mu}^T \mathbb{D}^T \dot{z} \\ &= -k_G \dot{z}^T \frac{\partial^2 f}{\partial z^2} \dot{z} = -\dot{z}^T H(z) \dot{z} \leq 0 \quad \forall \dot{z}, \dot{\mu} \end{aligned} \quad (14)$$

where $H(z) = k_G \frac{\partial^2 f}{\partial z^2} \succ 0$ due to strict convexity given by Assumption 1. As there is no dependence upon $\dot{\mu}$ in \dot{V} , LaSalle's invariance principle must be used to show convergence to $(\dot{z}, \dot{\mu}) = (0, 0)$.

Let the set where $\dot{V} = 0$ be denoted as

$$S = \{(\dot{z}, \dot{\mu}) | \dot{V} = 0\} = \{(\dot{z} = 0, \dot{\mu} \in \mathbb{R}^{Mn})\} \quad (15)$$

To see that that the only solution in which the complete state $(\dot{z}, \dot{\mu})$ can remain in S is the equilibrium $(0, 0)$, use the fact that to stay in $S \Rightarrow \dot{z} = 0 \quad \forall t \Rightarrow \ddot{z} = 0$. From this we see that

$$\ddot{z} = -H(z)\dot{z} - k'_I \mathbb{D} \dot{\mu} = -k_I \mathbb{D} \mathbb{D}^T z = -k_I \mathbb{L} z = 0,$$

where $k_I'^2 = k_I$. For the connected graph, the only z such that $-\mathbb{L}z = 0$ is $z = \alpha \otimes \mathbf{1}$, $\alpha \in \mathbb{R}^n$. This shows two things:

- (1) $x_i = x_j \quad \forall i, j$ which means that the agents reach consensus.
- (2) $\dot{\mu} = k'_I \mathbb{D}^T (\alpha \otimes \mathbf{1}) = 0$ which shows that the only possible value for $\dot{\mu}$ which stays in S is

$$\dot{\mu} = 0.$$

Since V is radially unbounded, this completes the proof. \square

Theorem 2. *Given Assumptions 1, 2, and 3 as well as the dynamics in (12) and (13), the saddle point $(\dot{z}, \dot{\mu}) = (0, 0)$ corresponds to the global minimum.*

Proof. To validate that a feasible solution is a local extremum, z^* , of a constrained optimization problem it is sufficient to show that z^* corresponds to a regular point (i.e. rows of $\frac{\partial h}{\partial z}(z^*)$ are linearly independent) and there exists λ^* such that

$$0 = \frac{\partial f}{\partial z}(z^*) + \lambda^{*T} \frac{\partial h}{\partial z} \quad (16)$$

where $h(z) = 0$ is the constraint and $f(z)$ is the cost, (see (Luenberger & Ye, 2008) for a discussion on local extremum and regular points). Due to Assumption 1, the only extremum is the global minimum. Therefore, this proof is performed in two steps. First, we show that the saddle point corresponds to a feasible point satisfying (16), then we show that the saddle point is indeed a regular point.

The proof of Theorem 1 showed that $(\dot{z}, \dot{\mu}) = (0, 0)$ implies that consensus is reached. Thus, the constraints are satisfied and the saddle point is feasible. Also, by noting that $\frac{\partial h}{\partial z} = \mathbb{D}^T$ for the problem at hand, (12) gives us

$$\dot{z} = 0 = -k_G \frac{\partial f}{\partial z} - k_I' \frac{\partial h}{\partial z} \mu \Rightarrow 0 = k_G \frac{\partial f}{\partial z} + k_I' \mu^T \frac{\partial h}{\partial z}. \quad (17)$$

Allowing $\lambda = \frac{k_I'}{k_G} \mu$, (16) is satisfied.

The saddle point must now be shown to be a regular point. To do so, we show that the convergent point is a regular point to the problem in which edges are removed from \mathcal{G} to form a minimum spanning tree (for undirected graphs, a minimum spanning tree is a connected graph with N nodes and $N - 1$ edges, e.g. (Mesbahi & Egerstedt, 2010)). Due to Assumption 3, a minimum spanning tree, \mathcal{G}_T , exists such that $\mathcal{E}_T \subset \mathcal{E}$. The saddle point is shown to be regular by first showing that the representation of the constraints using \mathcal{G}_T , i.e. $\mathbb{D}_T^T z = 0$, is linearly independent and then showing that if a λ can be found to satisfy (16) for \mathcal{G} , a λ_T can be found to satisfy (16) for \mathcal{G}_T .

Let $D_T \in \mathbb{R}^{N \times N-1}$ be the incidence matrix associated with \mathcal{G}_T . The graph Laplacian for a connected graph with N nodes always has rank $N - 1$, (Mesbahi & Egerstedt, 2010). Therefore, \mathbb{D}_T has full rank, which for $n = 1$, gives that $D_T^T z = 0$ is a linearly independent set of constraints. For

$n > 1$, $\mathbb{D}_T = I_n \otimes D_T$, and, as noted in Section 2.2, $\mathbb{D}_T^T z = \begin{bmatrix} D_T^T z_1 \\ \vdots \\ D_T^T z_n \end{bmatrix}$ which will also be linearly independent.

Without loss of generality, we can assume that $D = \begin{bmatrix} D_T & D_R \end{bmatrix}$ where D_R contains the “redundant” edges not contained in \mathcal{G}_T . Since D_T has the same rank as D , the columns in D_R can be expressed as linear combinations of the columns of D_T . In other words, $D_R = D_T \delta$, where $\delta \in \mathbb{R}^{N-1 \times M-N+1}$.

Without loss of generality, assume the elements in z have been rearranged to write $D = \begin{bmatrix} D_T & D_R \end{bmatrix}$, where $\mathbb{D}_R = I_n \otimes D_R$. Since $D_R = D_T \delta$, \mathbb{D}_R can be expressed as $\mathbb{D}_T \Delta$, where $\Delta = \mathbf{1} \otimes \delta$. We can separate λ as $\lambda = \begin{bmatrix} \lambda' \\ \lambda'' \end{bmatrix}$ which allows us to write $\mathbb{D}\lambda = \mathbb{D}_T \lambda' + \mathbb{D}_R \lambda'' = \mathbb{D}_T \lambda' + \mathbb{D}_T \Delta \lambda''$. Therefore, if a λ is found such that (16) is satisfied for \mathcal{G} , λ_T can be defined as $\lambda_T = \lambda' + \Delta \lambda''$. Thus, the solution is a regular point for the constraint $\mathbb{D}_T^T z = 0$. \square

3.2 Integral control

With the optimization framework in hand, the loop can be closed on the discussion begun in Section 2.3 by relating the dynamics in (12) and (13) to integral control. Note that the Lagrange multiplier, μ , can be expressed as follows (assuming $\mu(t_0) = 0$):

$$\mu(t) = \int_{t_0}^t \dot{\mu}(\tau) d\tau = \int_{t_0}^t k'_I \mathbb{D}^T z(\tau) d\tau = k'_I \mathbb{D}^T \int_{t_0}^t z(\tau) d\tau. \quad (18)$$

This allows \dot{z} to be expressed as:

$$\dot{z}(t) = -k_G \frac{\partial f^T}{\partial z} - k_I \mathbb{D} \mathbb{D}^T \int_{t_0}^t z(\tau) d\tau = -k_G \frac{\partial f^T}{\partial z} - k_I \mathbb{L} \int_{t_0}^t z(\tau) d\tau, \quad (19)$$

which gives the same result obtained in (8) assuming $k_P = 0$. After closer inspection of (13), one can see that the Lagrange multiplier, μ is indeed the integral of the weighted error referred to in Section 2.3.

3.3 Distributed implementation

While the analysis of this method has been performed from the point of view of the entire system, its utility as a distributed optimization technique would be questionable if it were not possible for the

algorithm to be executed by each agent using only local information. Therefore, we now present the algorithm in terms of implementation of a single agent and discuss the information and communication requirements.

Equations (12) and (13) can be written in terms of execution by a single agent, i , as follows:

$$\dot{x}_i = -k_G \frac{\partial f_i}{\partial x}^T(x_i) - k_I' \sum_{j \in \mathcal{N}_i} \mu_i^j, \quad (20)$$

$$\dot{\mu}_i^j = k_I'(x_i - x_j), \quad (21)$$

where for simplification we have introduced the Lagrange multiplier variables $\mu_i^j = -\mu_j^i = d_{i,k_{ij}} \mu_{k_{ij}}$ where k_{ij} is the edge connecting agents i and j and it is assumed that $\mu_k(0) = 0$, $k = 1, \dots, M$.¹ Note that \mathcal{N}_i denotes agent i 's neighborhood set, or agents with which agent i can communicate. By inspection, agent i can compute \dot{x}_i and $\dot{\mu}_i^j \forall j \in \mathcal{N}_i$ using only its own state and the states of its neighbors. Therefore, we emphasize that *the only piece of information that an agent needs to communicate with its neighbors is its version of the state vector*.

3.4 Example

To illustrate behaviors typical of dual decomposition, we give a numerical example. Let the individual costs be defined as follows:

$$\begin{aligned} f_1(x_1) &= (x_{11} - 1)^2 + \frac{1}{3}(x_{11} - x_{12})^2, \\ f_2(x_2) &= (x_{22} - 3)^2 + \frac{1}{3}(x_{21} - x_{22})^2, \\ f_3(x_3) &= (x_{31} - 6)^2 + \frac{1}{3}(x_{31} - x_{32})^2. \end{aligned} \quad (22)$$

where $x_i = \begin{bmatrix} x_{i1} & x_{i2} \end{bmatrix}^T$ and the network structure takes the form of the line graph shown in Figure 2. In other words, agents 1 and 2 as well as 2 and 3 can communicate, but agents 1 and 3 cannot. The global cost is given by $\sum_{i=1}^3 f_i(x_i)$, where $x_1 = x_2 = x_3$, has the optimal solution of $x^* = \begin{bmatrix} 3.4 & 3.2 \end{bmatrix}^T$.

Figure 3 and Table 1 show the results of employing these dynamics. As seen in Figure 3, there is oscillation in the solution as the different agents communicate and vary their values. This oscillation is quite typical of dual-decomposition (Rantzer, 2007), and it will be seen that the oscillation increases

¹By uniqueness of solutions to differential equations, $\dot{\mu}_i^j(t) = -\dot{\mu}_j^i(t) \forall t$



Figure 2: This figure depicts the “Line” network structure used for the examples in Sections 3, 4, and 5

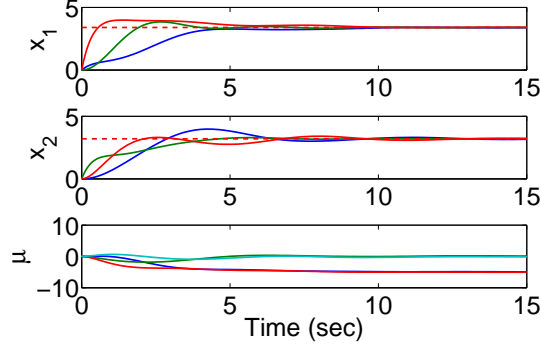


Figure 3: This figure shows the results from the convex optimization example using dual-decomposition

with an increase in problem complexity and number of agents in Section 6. Table 1 shows that the I control (corresponding to dual-decomposition) has a large overshoot and slower settling times when compared with the P and PI control laws (which are discussed in Sections 4 and 5). This is to be expected as the integral term will decrease the dampening of the system (Franklin et al., 2001). Moreover, as expected, Table 1 shows that there is zero steady-state error when using dual decomposition.

4 Consensus based distributed optimization

This section introduces the consensus-based distributed optimization technique, first outlined in (Nedic & Ozdaglar, 2009a), which will give the proportional component in the new PI distributed optimization method. After formulating the algorithm in terms of notation presented in previous sections, characteristics of the convergence are discussed in terms of the constrained optimization problem. This section will end by resuming the example started in the previous section to present a comparison between the distributed optimization methods.

4.1 Consensus based algorithm

While originally given in discrete time, we present the consensus based distributed optimization problem in continuous time as done in (Kvaternik & Pavel, 2012) to maintain consistent notation. In stark

contrast to the development of dual-decomposition, the consensus-based method was not designed from existing optimization methods. Rather, it was directly developed for networked, multi-agent systems. The foundation of this concept is that the consensus equation, a core equation in many multi-agent designs, e.g. (Mesbahi & Egerstedt, 2010; Olfati-Saber & Murray, 2004; Jadbabaie et al., 2003), can be used to force agreement between different agents. Therefore, the basic idea is for each agent to combine a step in the gradient direction with a step in the direction of consensus.

As the consensus method was developed for the multi-agent scenario, it can immediately be expressed in a distributed fashion as

$$\dot{x}_i = -k_G \frac{\partial f_i}{\partial x}(x_i) - \sum_{j \in \mathcal{N}_i} \alpha_{ij} (x_i - x_j), \quad (23)$$

where α_{ij} is the weighting that agent i associates with the edge of the graph connecting itself to agent j . Assuming equal weighting on all edges, i.e. $\alpha_{ij} = k_P \forall (v_i, v_j) \in \mathcal{E}$, the consensus based method can be stated for the aggregate state dynamics as:

$$\dot{z} = -k_P \mathbb{L}z - k_G \frac{\partial f^T}{\partial z}. \quad (24)$$

From this expression of the aggregate dynamics, we immediately see that the consensus term is the proportional term given in (8).

We do not present a proof of this method as it does not add to the development in this paper. For the discrete-time analog to (24), using a diminishing or adaptive step-size rule² for determining k_G at each iteration of the optimization would cause the agents to converge to the optimal value. For the continuous case, (Kvaternik & Pavel, 2012) proves that agents can come arbitrarily close to the optimum by choosing $\frac{k_G}{k_P}$ to be “sufficiently small.”

The diminishing step-size condition has been observed to be a possible deterrent of quick convergence of the algorithm, e.g. (Wang & Elia, 2010; Nedic & Ozdaglar, 2009a; Wang & Elia, 2011). To balance a tradeoff between convergence and optimality, (Nedic & Ozdaglar, 2009a) proposed a scheme of changing k_G during execution to get closer to the optimal point. The basic idea is that a constant gain often will result in the state approaching a steady-state value in relatively few steps. Once the state is “close enough” to the steady-state then the gain is changed to zero to allow the agents to reach consensus. They prove that the longer the agents wait to switch to the zero gain, the closer

²Section 4 is the only section which considers the gain k_G to be time-varying. Throughout the rest of the paper, all gains (k_G , k_I , and k_P) are considered constant.

they will come to the optimal value, but will suffer in convergence rate.

4.2 Consensus method and constrained optimization

We now examine this tradeoff further in terms of the underlying constrained optimization problem given in Section 2.3. This will give insight into the effect of the contribution of the proportional term and the benefit of including an integral term, which is done in Section 5.

To perform this analysis, assume that \bar{z} is the steady-state result of executing (24) as $t \rightarrow \infty$. Such a \bar{z} is known to exist due to the analysis in (Kvaternik & Pavel, 2012). At \bar{z} , (24) will give

$$\dot{z} = 0 = -k_P \mathbb{L} \bar{z} - k_G \frac{\partial f^T}{\partial z}(\bar{z}) \Rightarrow 0 = \frac{k_P}{k_G} \mathbb{L} \bar{z} + \frac{\partial f^T}{\partial z}(\bar{z}). \quad (25)$$

Using the fact that $\mathbb{L} = \mathbb{L}^T = \mathbb{D} \mathbb{D}^T$, (25) can be expressed as $0 = \frac{\partial f}{\partial z} + \frac{k_P}{k_G} \bar{z}^T \mathbb{D} \mathbb{D}^T$. Now, let $\lambda^T = \frac{k_P}{k_G} \bar{z}^T \mathbb{D}$ and recall that $\frac{\partial h}{\partial z} = \mathbb{D}^T$, where $h(z) = 0$ is the equality constraint. This gives $0 = \frac{\partial f}{\partial z} + \lambda^T \frac{\partial h}{\partial z}$ as in (16). While this satisfies part of the condition for determining an extreme point, \bar{z} will not be optimal as consensus will not be reached, resulting in the constraints not being met, (Kvaternik & Pavel, 2012).

As discussed in Section 2.2, $\mathbb{L}z$ will always point along lines perpendicular to the constraint set. This means that \bar{z} will be a point where $\frac{\partial f}{\partial z}(\bar{z})$ points along a line perpendicular to the constraint set. Now, let $\bar{z}' = z(t)$ as $t \rightarrow \infty$ where $\dot{z} = -\mathbb{L}z$ and $z(0) = \bar{z}$. Since $\mathbb{L}z$ points directly to the constraint set, \bar{z}' will be the point of intersection of the constraint set orthogonal to \bar{z} . Therefore, if $f(z)$ is such that the gradient will always point directly at the unconstrained optimal point, then the result of the optimization strategy proposed in (Nedic & Ozdaglar, 2009a) can converge arbitrarily close to the optimal value. An example of such a convex function is shown in Figure 1.

More important to our discussion is that a constantly weighted consensus term will not have enough control authority to pull the state of the system all of the way to the optimal point. However, it will help to guide the state to, and maintain it on, a line in which the only additional control effort need be in the direction of consensus. This further motivates the choice of adding an integral control term.

4.3 Example

We continue the example started in Section 3 using the consensus-based distributed optimization. Two scenarios are shown for the gain: $k_G = 1$ which violates the diminishing or adaptive gain requirement

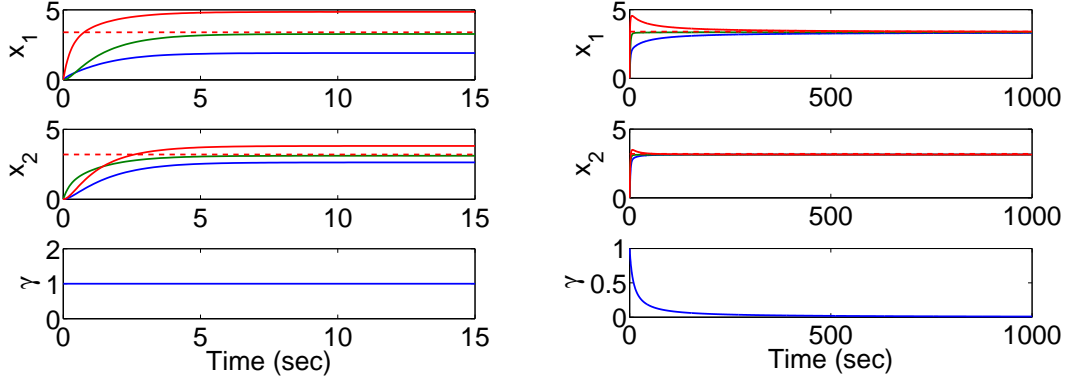


Figure 4: This figure shows the result of optimizing using consensus for the problem given in (22) for both a constant and fading value for k_G on the left and right respectively

and $k_G = \frac{1}{1+1t}$ which satisfies the requirement. The results are shown in Figure 4 and Table 1. The constant gain example exhibits the very desirable attribute of quick convergence, however suffers in performance as the values do not converge and the optimal value is not reached. On the other hand, the fading gain example shows that the optimal values can be achieved, but convergence suffers as expected. Both exhibit the desirable attribute of very little oscillation in the solution, however, the fading gain does show a significant increase in overshoot.

Remark 2. *In presenting examples throughout the remainder of the paper, the results from both a constant and a diminishing gain will be shown. We do this instead of trying to tune the “stopping” criteria given in (Nedic & Ozdaglar, 2009a). The result of a constant gain will emphasize the possible convergence rate and a diminishing gain will emphasize the ability to reach optimality.*

5 PI distributed optimization

In Sections 3 and 4, dual decomposition and the consensus method for distributed optimization were introduced and the parallel to integral and proportional control laws was seen. In this section, we show that these two methods can be combined to create a new distributed optimization method which is guaranteed to converge to the global minimum, much like integral control can be added to proportional control to achieve zero steady-state error with good convergence properties.

This section begins by developing the PI distributed optimization method and proving that it converges to the global minimum. The relationship to PI control is then discussed and the example of the previous two sections is finished.

5.1 PI distributed optimization algorithm

The PI distributed optimization algorithm is formed by noting that the dual-decomposition method discussed in Section 3 shares similar structure with the consensus method discussed in Section 4. Each has a gradient term along with an additional term added to enforce equality between agents. Dual-decomposition guarantees convergence to the goal, but has an undesirable transient, oscillatory behavior. On the other hand, the consensus method does not converge under constant gains, but has a much more damped transient response. Therefore, we join the two methods in a desire to achieve the benefits of each.

Combining equations (12) and (13) with (24), the aggregate dynamics can be expressed as

$$\begin{aligned}\dot{z} &= -k_G \frac{\partial f^T}{\partial z} - k_P \mathbb{L}z - k'_I \mathbb{D}\mu \\ \dot{\mu} &= k'_I \mathbb{D}^T z.\end{aligned}\tag{26}$$

Similarly, (20) and (21) can be combined with (23) to get a distributed implementation as follows:

$$\dot{x}_i = -k_G \frac{\partial f_i^T}{\partial x}(x_i) - k_P \sum_{j \in \mathcal{N}_i} (x_i - x_j) - k'_I \sum_{j \in \mathcal{N}_i} \mu_i^j.\tag{27}$$

$$\dot{\mu}_i^j = k'_I (x_i - x_j)\tag{28}$$

where we again define μ_i^j as in (21). As in Sections 3 and 4, *the only information exchange required between agents is the exchange of the state vectors between neighboring agents.*

To show convergence to the global minimum, we give the following two theorems.

Theorem 3. *Given Assumptions 1, 2, and 3 as well as the dynamics in (26), the saddle point $(\dot{z}, \dot{\mu}) = (0, 0)$ is globally asymptotically stable.*

Proof. The same proof can be used as was used in Theorem 1 with two modifications.

$$(1) \ H(z) = k_G \frac{\partial^2 f}{\partial z^2} + k_P \mathbb{L}, \text{ but } H(z) \succ 0 \text{ still holds.}$$

$$(2) \ \ddot{z} = -k_G \frac{\partial^2 f}{\partial z^2} \dot{z} - k_P \mathbb{L} \dot{z} - k'_I \mathbb{D} \dot{\mu} \text{ which when } \dot{z} = 0 \text{ still simplifies to } \ddot{z} = -k'_I \mathbb{D} \dot{\mu}$$

□

Theorem 4. *Given Assumptions 1, 2, and 3 as well as the dynamics in (26), the saddle point $(\dot{z}, \dot{\mu}) = (0, 0)$ corresponds to the global minimum.*

Proof. The same proof can be used as was used in Theorem 2 by noting for a feasible solution, $\mathbb{L}z = 0$. This will give the same equation for \dot{z} as given in (17). \square

The proofs of Theorems 3 and 4 basically show that adding the consensus term does not break the convergence properties of the dual-decomposition method of Section 3, but do nothing to speak of the benefit of adding the consensus term. To see the benefit of the consensus term, consider the following problem:

$$\begin{aligned} \min_z \quad & k_G f(z) + \frac{k_P}{2} z^T \mathbb{L} z. \\ \text{s.t.} \quad & k'_I \mathbb{D}^T z = 0 \end{aligned} \tag{29}$$

This is the same problem as given in (5), but with the addition of a term proportional to the square of the constraint (recall $\mathbb{D}\mathbb{D}^T = \mathbb{L}$). Adding the square of the constraint is known as the augmented Lagrangian method, which has been shown to add dampening to the dual optimization problem, improving convergence, (see (Boyd, Parikh, Chu, Peleato, & Eckstein, 2011) for a discussion and analysis of the augmented Lagrangian).

Following the same method to develop dynamic update laws as in Section 3, the following dual optimization problem would be solved:

$$\max_{\mu} \min_z \left(k_G f(z) + k'_I z^T \mathbb{D} \mu + \frac{k_P}{2} z^T \mathbb{L} z \right), \tag{30}$$

with the resulting dynamics being the same as (26). Thus, adding in a consensus term corresponds to modifying the problem to solve the augmented Lagrangian, producing the desired dampening effect without modifying the guarantee of convergence.

5.2 Connections to PI control

As with the previous two distributed optimization techniques, we note the similarity of this distributed optimization framework with a PI control framework. The Lagrange multiplier, μ , can be expressed in the same form as done in (18). Thus, the following expression for \dot{z} can be obtained:

$$\dot{z}(t) = -k_G \frac{\partial f^T}{\partial z} - k_I \mathbb{L} \int_{t_0}^t z(\tau) d\tau - k_P \mathbb{L} z(t). \tag{31}$$

This is the same equation that was derived for a PI control law in Section 2.3. We can therefore expect to see properties of PI control such as increased overshoot resulting from decreased dampening of the

proportional control, zero steady-state error due to the integral term (which has already been proved), and faster settling time than pure integral control, e.g. (Franklin et al., 2001).

While there exist many distributed optimization techniques, e.g. (Palomar & Eldar, 2010; Wei & Ozdaglar, 2012; Kvaternik & Pavel, 2012; Wang & Elia, 2010, 2011; Terelius et al., 2011), it is important to note the similarity of the method in this section to that presented in (Wang & Elia, 2010) and extended in (Wang & Elia, 2011; Gharesifard & Cortés, 2012). While the development of the algorithm in (Wang & Elia, 2010) is different than the development in this paper, it can be expressed as using the augmented Lagrangian to solve the following problem:

$$\min_z f(z), \quad (32)$$

$$\text{s.t. } \mathbb{L}z = 0$$

where the resulting dynamics can be expressed as

$$\dot{z} = -\frac{\partial f^T}{\partial z}(z(t)) - \mathbb{L}z(t) - \mathbb{L}\mu(t), \quad (33)$$

$$\dot{\mu} = \mathbb{L}z, \quad (34)$$

and now $\mu \in \mathbb{R}^{Nn}$ as opposed to $\mu \in \mathbb{R}^{Mn}$ as before. The only difference between this method and the method we have developed is simply that the constraint is expressed in terms of the graph Laplacian instead of the incidence matrix. This would result in an equation similar to (31), except with an \mathbb{L}^2 term instead of an \mathbb{L} term in front of the integral.

While this may seem like a small difference, due to the fact that we have utilized dual-decomposition in the development of the integral term, *we form a PI distributed optimization technique which requires half of the communication* that the technique developed in (Wang & Elia, 2010) requires. This can be seen from the fact that the incidence matrix, used in dual-decomposition, allows each agent to update the necessary values of μ using only local information. However, using the Laplacian matrix to express the constraint forms an \mathbb{L}^2 term which requires that either each agent knows their neighbors' neighbors states or each neighbor must additionally communicate μ_i at each optimization step.

5.3 Example

We continue the example in (22) using the newly derived dynamics. In Figure 5, it is apparent that the PI optimization is able to achieve zero error while converging quickly and with little oscillation.

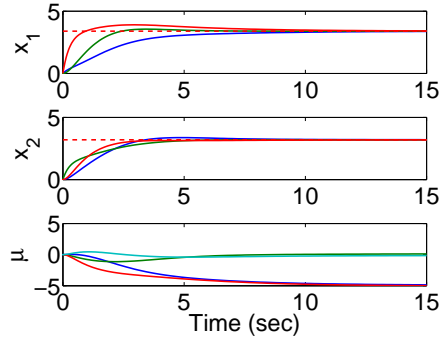


Figure 5: This figure shows the results from the convex optimization example using PI distributed optimization

Furthermore, Table 1 shows that settling time and overshoot are in between the values of pure proportional and pure integral control, as expected. These attributes will be emphasized in the examples in the following sections as more complex problems are presented.

6 Scalable multi-agent formulation

Up until this point, we have presented the algorithms in terms of a framework where each agent keeps its own version of the entire state vector as done in previous works, e.g. (Kvaternik & Pavel, 2012; Wang & Elia, 2010; Nedic & Ozdaglar, 2009a; Terelius et al., 2011). This is not necessary if some of the agents' individual costs do not depend upon all of the elements of the parameter vector being optimized. An example of this will be shown at the end of the section where each agent introduces more parameters to be optimized, typical in multi-robot scenarios, e.g. (Keviczky et al., 2006; Droge & Egerstedt, 2013; Rantzer, 2009; Dunbar & Murray, 2006). However, each agents' cost depends solely on the parameters introduced by its neighbors. In such a situation, it is not necessary for each agent to keep track of the entire parameter vector and, in fact, doing so is not scalable to large numbers of agents.

In this section, we address this in a similar fashion to (Wang & Elia, 2011) and show that it fits quite naturally into the framework of the previous sections. First, it is shown that even with the reduction of parameters the previous theorems still hold. Then, the reduction of parameters will lead to a slight reformulation of the PI distributed optimization algorithm. Finally, we end this section with an example where drastic improvement in convergence is achieved by reducing the number of variables that each agent must maintain.

6.1 Eliminating unneeded variables

When each agent does not have an opinion about a parameter in the parameter vector, the problem can be simplified to eliminate redundancies. Similar to (Wang & Elia, 2011), let $I_j = \{i | f_i \text{ depends on the element } j\}$ be the set of agents which depend on element j with cardinality $N_j = |I_j|$. As agents no longer need to keep track of the entire vector, the definition of z_j needs to be slightly modified to $z_j \triangleq \text{vec}[x_{ij}]_{i \in I_j} \in \mathbb{R}^{N_i}$, a subset of the elements originally contained in z_j . Now, the aggregate vector can be defined as $z = \begin{bmatrix} z_1^T & \dots & z_n^T \end{bmatrix}^T \in \mathbb{R}^{N_1 + \dots + N_n}$.

Let the induced subgraphs, $\mathcal{G}_i(\mathcal{V}_i, \mathcal{E}_i)$, be defined as $\mathcal{V}_i = \{v_j \in \mathcal{V} | j \in I_i\} \subseteq \mathcal{V}$ and $\mathcal{E}_i = \{(v_i, v_j) \in \mathcal{E} | v_i, v_j \in \mathcal{V}_i\}$. Finally, the following assumption is made to allow for convergence

Assumption 4. \mathcal{G}_i is connected $\forall i \in \{1, \dots, N\}$.

Note that, given Assumption 3, Assumption 4 is not limiting. If there exists i s.t. \mathcal{G}_i is not connected, one needs only to extend \mathcal{G}_i to contain nodes originally in \mathcal{G} that will connect the different connected components of \mathcal{G}_i .

Along this same line of reasoning, we briefly touch upon a topic of study which is out of the scope of this paper, but worth mentioning. There may be simple cases in which choosing \mathcal{G}_i such that it is connected with the smallest number of vertices possible will not result in the fastest convergence to the global minimum. There has been much work done on the convergence of the consensus equation and the network topology plays a key role in determining the convergence rate (Mesbahi & Egerstedt, 2010; Olfati-Saber & Murray, 2004; Jadbabaie et al., 2003). Therefore, to achieve the fastest performance, selection of the sub-graph for each variable could be more complicated than simply choosing the minimally connected sub-graph.

In any case, given \mathcal{G}_i , the corresponding incidence matrix, $D_i \in \mathbb{R}^{N_i \times M_i}$, where $M_i = |\mathcal{E}_i|$, and graph Laplacian, $L_i \in \mathbb{R}^{N_i \times N_i}$, can be defined. This allows for the definition of the aggregate matrices $\mathbb{D} \triangleq \text{diag}(D_1, \dots, D_n)$ and $\mathbb{L} = \text{diag}(L_1, \dots, L_n)$. These aggregate matrices will continue to exhibit the same properties mentioned in Section 2.2 as they can still be expressed as n connected components of a graph. The only difference is that the connected components do not have the same structure. As these properties still hold, Theorems 1 through 4 will also hold using the newly defined augmented matrices and addition of Assumption 4.

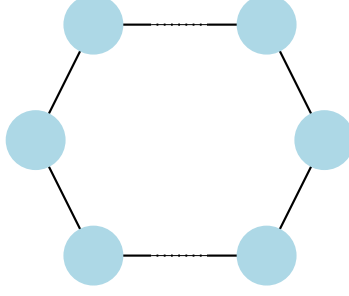


Figure 6: This figure depicts the “Ring” network structure used in Section 6

6.2 Distributed implementation

While the aggregate dynamics of the multi-agent system can be expressed without any change, the dynamics executed by each agent will change slightly due to the fact that each variable in the parameter vector will have a different set of agents that are maintaining a version of it. We express the dynamics of a single variable as follows:

$$\dot{x}_{ij} = -k_G \frac{\partial f_i}{\partial x_{ij}} - k_P \sum_{k \in \{\mathcal{N}_i \cap I_j\}} (x_{ij} - x_{kj}) - k'_I \sum_{k \in \{\mathcal{N}_i \cap I_j\}} \mu_{i,j}^k \quad (35)$$

$$\dot{\mu}_{i,j}^k = k'_I (x_{ij} - x_{kj}). \quad (36)$$

Note that the algorithms in Sections 3 and 4 can be achieved by setting $k_P = 0$ and $k'_I = 0$ respectively. Again, we see that *each agent is able to execute this algorithm using local information and only communicating its version of the parameters being optimized with its neighbors.*

6.3 Ring example

We now present an example in which scaling down the number of parameters that each agent worries about drastically improves the performance of the system. Consider the “Ring” network depicted in Figure 6 where each agent can communicate with agents to each side. In this example, each agent has a variable that “belongs” to it and it wants to balance having its value be close to its neighbors’ value as well as a nominal value. This can be expressed in the form of the following quadratic cost:

$$f_i = (x_{i,i-1} - x_{ii})^2 + (x_{ii} - x_{d_i})^2 + (x_{ii} - x_{i,i+1})^2 \quad (37)$$

where $x_{d_i} = i$ is the desired value.

Note that for the formulation in Sections 3, 4, and 5, each agent would have had to keep track of $N = 20$ variables, corresponding to the aggregate state vector having 400 elements. However, this is greatly reduced by following the formulation in this section. Each agent will only need to keep track of 3 variables with a total of 60 variables in the aggregate state vector.

The results of both representations of the state can be seen in Figure 7 and Tables 2 and 3. Significant improvement can be seen across the board in terms of settling time for reducing the number of variables. Moreover, the overshoot is drastically improved for both the I and PI distributed optimization methods. Related to overshoot, it is seen in Figure 7 that the oscillation is drastically reduced for dual-decomposition.

One final observation about the performance of the PI distributed optimization technique is noteworthy. This example demonstrates the performance of the system when a larger number of variables is in question. We see in Table 2 that the PI distributed optimization significantly outperforms the other methods in terms of convergence. Moreover, there is a drastic improvement over the dual-decomposition method in terms of overshoot and oscillation as well as an improvement over the consensus method in terms of steady-state error.

Again, we emphasize that this is an extreme example meant to demonstrate the possible utility of reducing the number of variables that each agent deals with. Conclusions should not be drawn beyond the notion that this may be beneficial as there may be instances in which scaling back as much as possible would not be beneficial.

7 Conclusion

We have developed a new, PI distributed optimization method through the combination of dual decomposition and the consensus method for distributed optimization. This has been done by noting the similarity of the methods when considering the underlying constrained optimization problem. This new method is able to achieve desirable properties from both of the previous methods. Namely, faster convergence and dampening due to the proportional term, originating from the consensus based method, and zero steady-state error from the integral term, originating from dual-decomposition. The method was also modified to allow agents to maintain only the variables they care about, with an example showing drastic improvement in convergence times.

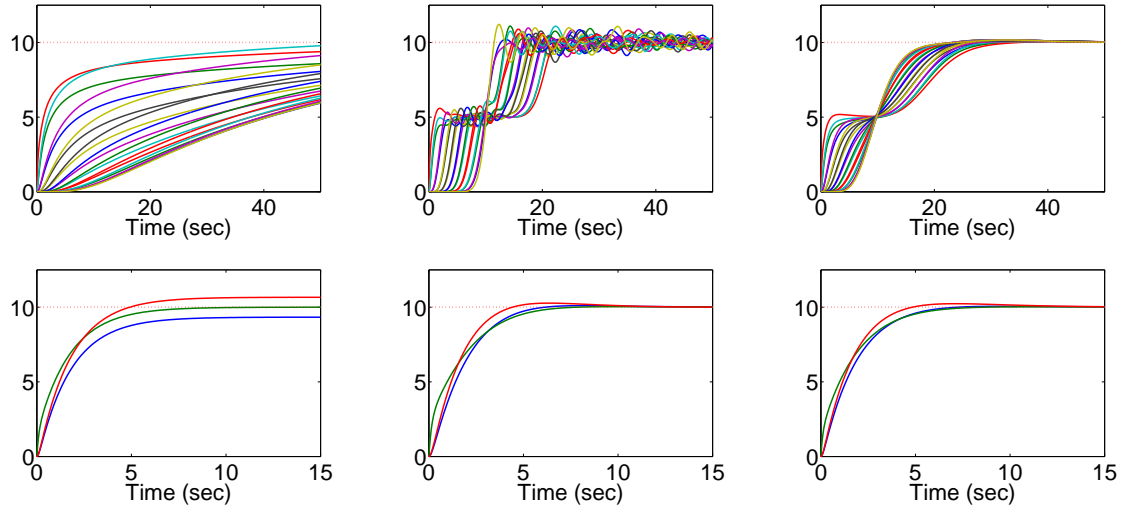


Figure 7: This figures shows the results of applying the formulation of Sections 3, 4, and 5 on the top row and 6 bottom row to solve the problem in (37). The left, middle, and right images of each row correspond to consensus, dual-decomposition, and PI distributed optimization techniques. The results shown are for variable 10. The solutions in the top row require 20 versions of this variable to converge to the optimal value where the solutions in the bottom row require only 3.

Funding

The work by M. Egerstedt was funded by The Air Force Office of Scientific Research through grant number [2012-00305-01]

References

- Arrow, K., Hurwicz, L., & Uzawa, H. (1958). Studies in nonlinear programming. Stanford University Press, Stanford, CA.
- Boyd, S. (2004). Convex optimization. Cambridge: Cambridge University Press.
- Boyd, S., Parikh, N., Chu, E., Peleato, B., & Eckstein, J. (2011). Distributed optimization and statistical learning via the alternating direction method of multipliers. Foundations and Trends in Machine Learning, 3(1), 1–122.
- Droge, G., & Egerstedt, M. (2013). Distributed parameterized model predictive control of networked multi-agent systems. American Control Conference.
- Dunbar, W., & Murray, R. (2006). Distributed receding horizon control for multi-vehicle formation stabilization. Automatica, 42(4), 549–558.
- Feijer, D., & Paganini, F. (2010). Stability of primal-dual gradient dynamics and applications to network optimization. Automatica, 46(12), 1974–1981.
- Franklin, G., Powell, J., & Emami-Naeini, A. (2001). Feedback control of dynamic systems. Prentice Hall.
- Gharesifard, B., & Cortés, J. (2012). Continuous-time distributed convex optimization on weight-balanced digraphs. In Ieee 51st annual conference on decision and control (cdc) (pp. 7451–7456).

- Giselsson, P., & Rantzer, A. (2010). Distributed model predictive control with suboptimality and stability guarantees. In 49th IEEE conference on decision and control (pp. 7272–7277).
- Jadbabaie, A., Lin, J., & Morse, A. (2003). Coordination of groups of mobile autonomous agents using nearest neighbor rules. IEEE Transactions on Automatic Control, 48(6), 988–1001.
- Keviczky, T., Borrelli, F., & Balas, G. (2006). Decentralized receding horizon control for large scale dynamically decoupled systems. Automatica, 42(12), 2105–2115.
- Kvaternik, K., & Pavel, L. (2011). Lyapunov analysis of a distributed optimization scheme. In 5th international conference on network games, control and optimization (netgcoop).
- Kvaternik, K., & Pavel, L. (2012). A continuous-time decentralized optimization scheme with positivity constraints. In IEEE conference on decision and control.
- Lobel, I., & Ozdaglar, A. (2008). Distributed subgradient methods over random networks. In Proceedings of allerton conference on communication, control, computation.
- Luenberger, D., & Ye, Y. (2008). Linear and nonlinear programming (3rd ed.) (Vol. 116). Springer.
- Mesbahi, M., & Egerstedt, M. (2010). Graph theoretic methods in multiagent networks. Princeton University Press.
- Nedic, A., & Ozdaglar, A. (2009a). Distributed subgradient methods for multi-agent optimization. IEEE Transactions on Automatic Control, 54(1), 48–61.
- Nedic, A., & Ozdaglar, A. (2009b). Subgradient methods for saddle-point problems. Journal of optimization theory and applications, 142(1), 205–228.
- Olfati-Saber, R. (2007). Distributed kalman filtering for sensor networks. In 46th IEEE conference on decision and control.
- Olfati-Saber, R., & Murray, R. (2004). Consensus problems in networks of agents with switching topology and time-delays. IEEE Transactions on Automatic Control, 49(9), 1520–1533.
- Palomar, D., & Eldar, Y. (2010). Convex optimization in signal processing and communications. Cambridge University Press.
- Rantzer, A. (2007). On prize mechanisms in linear quadratic team theory. In 46th IEEE conference on decision and control (pp. 1112–1116).
- Rantzer, A. (2009). Dynamic dual decomposition for distributed control. In American control conference (pp. 884–888).
- Shamma, J. (2007). Cooperative control of distributed multi-agent systems. Wiley Online Library.
- Sundhar Ram, S., Nedić, A., & Veeravalli, V. (2010). Distributed stochastic subgradient projection algorithms for convex optimization. Journal of optimization theory and applications, 147(3), 516–545.
- Terelius, H., Topcu, U., & Murray, R. (2011). Decentralized multi-agent optimization via dual decomposition. In Kth, automatic control. IFAC.
- Wang, J., & Elia, N. (2010). Control approach to distributed optimization. In 48th annual allerton conference on communication, control, and computing (allerton) (pp. 557–561).
- Wang, J., & Elia, N. (2011). A control perspective for centralized and distributed convex optimization. In 50th IEEE conference on decision and control and European control conference (cdc-ecc).
- Wei, E., & Ozdaglar, A. (2012). Distributed alternating direction method of multipliers. In IEEE conference on decision and control.

Table 1: The results of performing proportional, integral, and PI distributed optimization for the convex optimization problem

	P: $\gamma = 1$	P: $\gamma = \frac{1}{1+.1t}$	I	PI
M	0.11%	34.66%	24.24%	14.95%
t_{10}	3.54	103.73	5.61	5.14
t_1	6.66	869.32	15.04	13.19
% error	43.58%	1.97%	0%	0%

Table 2: The results of performing proportional, integral, and PI distributed optimization with each agent optimizing over the full state vector

	P: $\gamma = 1$	P: $\gamma = \frac{1}{1+.1t}$	I	PI
M	0.1%	0.12%	37.5%	7.9%
t_{10}	120.8	659.42	115.28	29.78
t_1	226.58	4884.8	542.71	83.02
% error	55.4%	0.92%	0%	0%

Table 3: The results of performing proportional, integral, and PI distributed optimization with each agent optimizing over a subset of the state vector

	P: $\gamma = 1$	P: $\gamma = \frac{1}{1+.1t}$	I	PI
M	0.1%	35.15%	7.12%	4.51%
t_{10}	5.2	82.85	6.12	6.03
t_1	9.47	692.57	12.78	12.33
% error	57.48%	5.3%	0%	0%

List of Figures

1	This figure shows the results of using the cost $f(z) = (x_1 - 1)^2 + (x_2 + 1)^2$. Left: Dotted line shows the equality constraint and the arrows show the gradient and projected gradient. Right: Result of performing the PI gradient method for optimization given in (8). The trajectory of the two states is shown ending in the final condition denoted by the solid circle and the constraint is shown as a dotted line. The arrows show the final gradient and Lagrange multiplier multiplied by the constraint. As expected, these are equal in magnitude, but opposite in direction.	8
2	This figure depicts the “Line” network structure used for the examples in Sections 3, 4, and 5	16
3	This figure shows the results from the convex optimization example using dual-decomposition	16
4	This figure shows the result of optimizing using consensus for the problem given in (22) for both a constant and fading value for k_G on the left and right respectively	19
5	This figure shows the results from the convex optimization example using PI distributed optimization	23
6	This figure depicts the “Ring” network structure used in Section 6	25
7	This figures shows the results of applying the formulation of Sections 3, 4, and 5 on the top row and 6 bottom row to solve the problem in (37). The left, middle, and right images of each row correspond to consensus, dual-decomposition, and PI distributed optimization techniques. The results shown are for variable 10. The solutions in the top row require 20 versions of this variable to converge to the optimal value where the solutions in the bottom row require only 3.	27

Notes on contributors



Greg Droge received the B.S. degree in Electrical Engineering from Brigham Young University, Provo, Utah. He received his M.S. and Ph.D. from the school of Electrical and Computer Engineering at the Georgia Institute of Technology in 2012 and 2014, respectively. His research interests include model predictive control and distributed optimization with emphasis in motion planning and control.



Hiroaki Kawashima received his M.S. and Ph.D. in informatics from Kyoto University, Japan in 2001 and 2007, respectively. He is currently a senior lecturer at the Graduate School of Informatics, Kyoto University, Japan. From 2010 to 2012, he was a JSPS Postdoctoral Fellow for Research Abroad, and a visiting researcher at the School of Electrical and Computer Engineering, Georgia Institute of Technology. His research interests include hybrid systems, networked control systems, pattern recognition, machine learning, and human-computer interaction. He is a member of the IEEE.



Magnus B. Egerstedt (S'99-M'00-SM'05-F'12) is a Professor in the School of Electrical and Computer Engineering at the Georgia Institute of Technology. He received the M.S. degree in Engineering Physics and the Ph.D. degree in Applied Mathematics from the Royal Institute of Technology, Stockholm, Sweden, in 1996 and 2000 respectively, and he received the B.A. degree in Philosophy from Stockholm University in 1996. Dr. Egerstedt's research interests include hybrid and networked control, with applications in motion planning, control, and coordination of mobile robots, and he is the director of the Georgia Robotics and Intelligent Systems Laboratory (GRITS Lab), is a Fellow of the IEEE, and received the CAREER Award from the U.S. National Science Foundation in 2003.